Title: How Do They Relate? An Investigation of Relationships That Appear Linear

Brief Overview:

This unit is designed to allow students to investigate real world data that follows a linear trend. Students will begin with introductory tasks, such as making crude predictions from a given set of data. The unit will end with more in–depth problem–solving, including using the graphing calculators to find equations for line of best fit, interpolating and extrapolating data. Summaries will include identifying when it is not appropriate to make predictions using a linear regression model.

NCTM Content Standard/National Science Education Standard:

Number and Operations
- Compute fluently and make reasonable estimates

Algebra
- Understand patterns, relations, and functions
- Represent and analyze mathematical situations and structures using algebraic symbols
- Use mathematical models to represent and understand quantitative relationships
- Analyze change in various contexts

Measurement
- Understand measurable attributes of objects and the units, systems, and processes of measurement
- Apply appropriate techniques, tools, and formulas to determine measurements

Data Analysis and Probability
- Select and use appropriate statistical methods to analyze data
- Develop and evaluate inferences and predictions that are based on data

Problem Solving
- Build new mathematical knowledge through problem solving
- Solve problems that arise in mathematics and in other contexts
- Apply and adapt a variety of appropriate strategies to solve problems
- Monitor and reflect on the process of mathematical problem solving

Communication
- Organize and consolidate their mathematical thinking through communication
- Communicate their mathematical thinking coherently and clearly to peers, teachers, and others
- Use the language of mathematics to express mathematical ideas precisely

Connections
- Recognize and use connections among mathematical ideas
- Understand how mathematical ideas interconnect and build on one another to produce a coherent whole

- Recognize and apply mathematics in contexts outside of mathematics

Grade/Level:

Grades 7 –10, Algebra I

Duration/Length:

Three 80 – 90 minute lessons

Student Outcomes:

Students will:

- Collect and store data using measurement and technology resources
- Represent data graphically
- Write the equation for and draw a trend line for given data
- Use technology (graphing calculator) to find the line of best fit
- Identify correlations within data (positive, negative, none)
- Interpret slope and *y* – intercept within the context of the problem
- Use the line of best fit to predict within the data set (interpolation)
- Use the line of best fit to predict outside the data set (extrapolation)
- Identify proper and improper use of statistics

Materials and Resources:

- Stopwatch
- Student Worksheets
- Teacher Notes
- Class set of graphing calculators
- Class set of measuring tape
- Worksheets:
    - Getting Ready for LinReg
    - Hand Pulse
    - Estimating Lines of Best Fit
    - Could the Shortest Man in the World See His Shadow?
    - At The Pump
    - To Create a Scatter Plot
    - To Find the Line of Best Fit
    - To Predict Values
    - Bring 'Em Home Albert
    - Olympic Gold Medals
    - Air Pollution and Asthma
    - Linear Regression Summative Assessment
    - Essay

Development/Procedures:

Lesson 1

Preassessment – The students will use the worksheet, "Getting Ready for LinReg" as a warm–up sheet. Students should be able to graph data points, graph a line with a given equation, and determine the equation of a given line as prerequisites to this unit.

Launch – Have two students stand in a circle starting, holding hands. Instruct one student at their seat to be the timer. When the teacher says go, and the timer will time how long it takes for the students to pass a "hand squeezing" pulse around the circle. When the first student squeezes their right hand, the second student will then squeeze their right hand and so forth until the pulse comes back to the first person. When it reaches the first person, that person will say stop and the student will stop timing and announce the time. Record the data on the front board, and instruct the students to record the data on the worksheet, "Hand Pulse". Continue to add students to the circle until the data chart is complete. Students can work up until the stop sign in the worksheet, and then wait for further instruction from the teacher.

Teacher Facilitation – Present the following information and vocabulary, to allow students to finish the "Hand Pulse" activity: trend line, line of best fit, positive correlation, negative correlation, no correlation, strong correlation, weak correlation, interpolation and extrapolation using a graph, interpolation and extrapolation using a linear equation, and linear regression. The teacher should let the students complete the "Hand Pulse" portion of the worksheet and fill in the chart. Then the teacher should model for the students the drawing of a trend line and explain positive, negative and no correlation. Then the teacher should model prediction using the trend line, and the estimation of the slope and y-intercept of the trend line. Teacher should then explain what each of them mean in context of the problem, and model how to predict using the equation.

Student Application – Assign the remaining questions in the "Hand Pulse" activity.

Embedded Assessment – During the entire lesson, the teacher should be actively circling the classroom, observing, probing, and guiding students with constructivist questions. Instead of answering a student's content questions, a teacher should reply with a question, whose answer will guide the student to find the answer on his or her own.

Reteaching/Extension –

- The class can regroup and discuss its findings together.
- If time remains, the teacher can use any form of cooperative learning (think–pair–share, matching, etc.) to ensure that students understand key concepts.
- Exit Ticket: Ask students to write a hypothetical set of data, such as values that may have occurred over the years 2001 to 2006. For example, a student may choose to pick "Bobby's Weight" or "Cost of a DVD Burner". Have students estimate whether the trend line would show a positive, negative, or no correlation and why.

Lesson 2

Preassessment – The students will use the "Estimating Lines of Best Fit" warm up worksheet in order to review how to find correlations and estimate lines of best fit from a scatter plot by hand.

Launch – Divide the class into small groups of two or three, and review the class rules and expectations for going outside for the day's lesson. Students will be measuring the height and shadow length of everyone in their group. Review the steps for this lab activity prior to the beginning of class. This lab is found on worksheet, "Could the Shortest Man in the World See his Shadow". Distribute tape measures and copies of this worksheet to students before going outside. Students should complete exercises 1 – 6 while outside, and return inside. Before the completion of question 7-17, the teacher should prompt student to review what they learned yesterday, including the steps of estimating a trend line. Then the students should complete questions 7-17.

Teacher Facilitation – Stop students at the end of exercise 17, and review with the students how to put data into the graphing calculator. Use the calculator to graph the data from "Could the Shortest Man in the World See his Shadow?" into a scatter plot. These steps can be found on the "To Create a Scatter Plot" handout. Walk students through how to calculate an accurate line of best fit, referring to the handout, "To Find a Line of Best Fit" for these steps. Calculate the line of best fit for the shadow data, and explain the meaning of the correlation coefficient means. Prompt students to explain how they would predict other values for shadow length when given a height value. Further prompt students to explain how they would predict other values for height when given a value for shadow length. Show students how to do this using the calculator. These steps can be found on the "To Predict Values" handout.

Student Application – The students will then complete questions 19–27 on the "Could the Shortest Man in the World See his Own Shadow" worksheet. The completed worksheet should be collected to assess understanding. Note, however, that an answer sheet has not been included because for most questions, the answers will vary based on student data collected.

Embedded Assessment – During the entire lesson, the teacher should be actively circling the classroom, observing, probing, and guiding students with constructivist questions. Instead of answering a student's content questions, a teacher should reply with a question, whose answer will guide the student to find the answer on his or her own.

Reteaching/Extension –

- Steps will be reviewed for how to calculate the line of best fit and predict values.
- Assign the application problem from worksheet, "At The Pump". This worksheet assesses student understanding of finding the line of best fit, by hand and on the calculator, predicting values, as well as the interpretation of parts of the equation.

Lesson 3

Preassessment – The students will use the "Bring 'Em Home Albert" warm up worksheet in order to review how to find associations and estimate lines of best fit from a table and scatter plot using the graphing calculator.

Launch –Allow students to work in pairs for a think–pair–share style to demonstrate application in real life situations of the last day's lessons. Students should come up with two real world examples for each of the following: positive correlation, negative correlation, and no correlation. Students should be prepared to explain why each would have the specified correlation.

Teacher Facilitation – The teacher should have students present their examples of positive, negative, and no correlation to the class. The reasons why each is a good example should be carefully outlined in the context of the variables. Then, the teacher should introduce to students how to work with data that uses year (i.e. 1980) as the independent variable. Explain to students that it is reasonable to have the lowest year be the starting year, and change that year to 0, and every year after that will be represented as "the

years away from" that starting year. This is how it is generally presented on the High School Assessment. Definitions of interpolation and extrapolation should also be discussed, stressing that the further away values sought are from the values known, the less accurate our predictions will be. Realism of negative and infinite predictions should be discussed. Also, before discussing the high jump worksheet, ask students to guess the best gold medal high jump height, 7'10". Mark this height with tape on the wall so that students can visualize the magnitude and difficulty of the height. Students could also run the 100m outside to see how their times compare to Olympic runners.

Student Application – The students will then complete the questions on the 100m Dash, High Jump and Air Pollution worksheets.

Embedded Assessment – During the entire lesson, the teacher should be actively circling the classroom, observing and guiding students with constructivist questions. The teacher should use the Socratic method to guide the student to the answers and conclusions.

Reteaching/Extension –

- Steps will be reviewed for how to calculate the line of best fit using the calculator.
- Investigate whether or not there is a correlation between height of an Olympic jumper and height jumped.

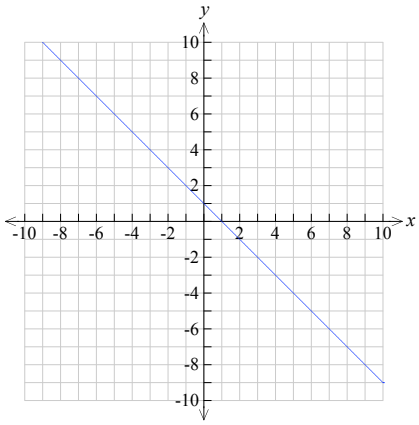
Summative Assessment:

Students will complete the attached Linear Regression summative assessment.

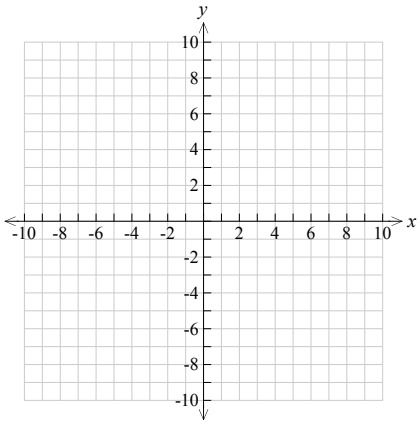
Getting Ready for LinReg

Name: _____
Date: _____

1. Write the equation of the following line, in slope–intercept form.

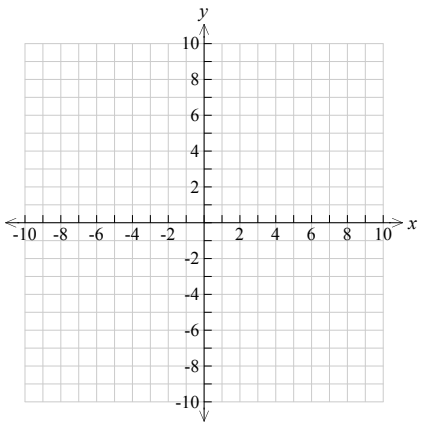


2. Graph the equation: $y = \frac{3}{5}x - 4$



3. Plot the points in the given table.

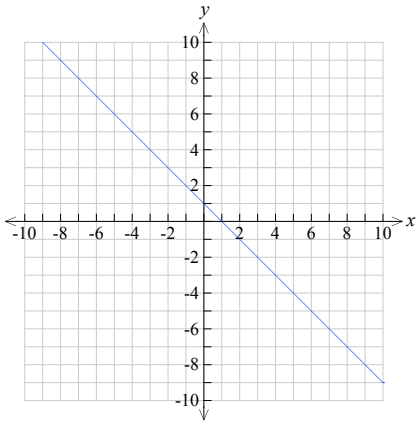
x	y
–4	4
–3	3
–2	2
–1	4
0	1
2	3
4	–1
5	0



Getting Ready for LinReg

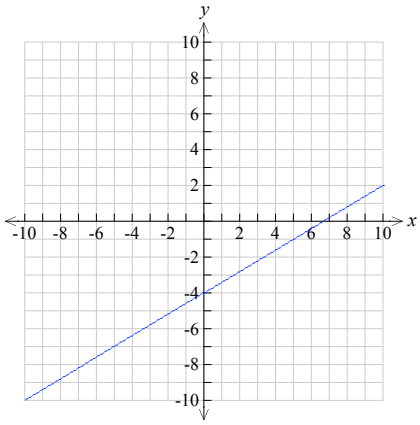
Name: __ANSWER KEY__
Date: _____

1. Write the equation of the following line, in slope–intercept form.



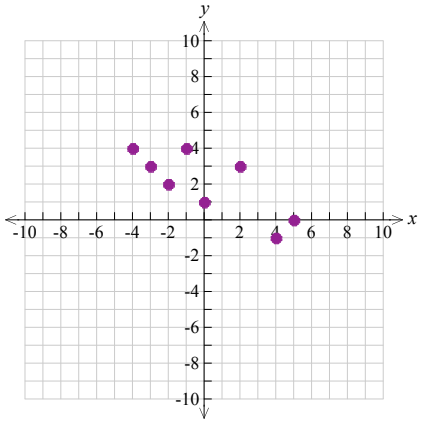
Answer:
$y = -x + 1$

2. Graph the equation: $y = \frac{3}{5}x - 4$



3. Plot the points in the given table.

x	y
–4	4
–3	3
–2	2
–1	4
0	1
2	3
4	–1
5	0



Hand Pulse

Name: _____
Date: _____

In this lab, we will measure the amount of time it takes for a 'hand pulse' to travel across the hands of specified numbers of students. We will learn then use the data to create a line of best fit, which can be used to make predictions about other data points which would likely occur. A line of best fit can be a very powerful tool for bringing data to life.

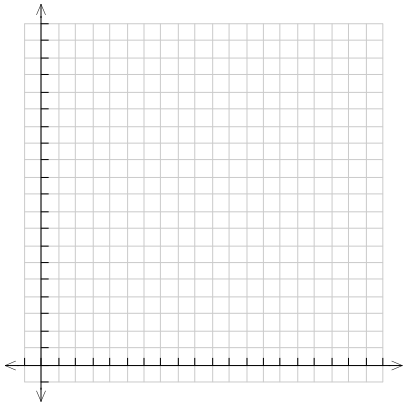
1. Carry out the hand–pulse activity. Record your data in the table below.

Number of Students	2	3	4	5	6	7	8	12	16
Time (s)									

2. What happens to the time as the number of students increases?

3. If the entire class were to try the hand–pulse activity, how long would you expect the pulse to take before it reaches the starting point? Explain.

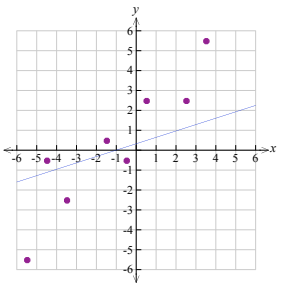
4. Graph your data below. Remember to title your graph and label your axes!



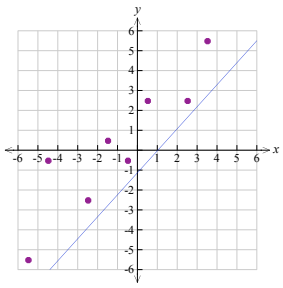
5. Looking at the pattern in the graph, how long would you expect the pulse to take to travel through the entire class? How much different was your answer, compared to #3?

6. A **trend** line is a line that best follows the 'trend' of the data. As best as possible, there should be an equal amount of points above and below the line.

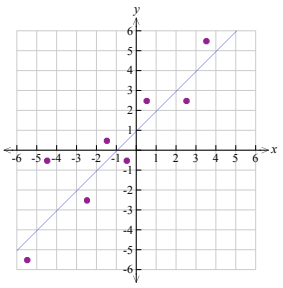
Examples of trend lines



This is not a good trend line. It does not follow the overall trend, or direction, of the data.



This is not a good trend line because the points are not equally balanced on both sides of the line.



This is a good trend line because the points are equally balanced on both sides of the line.

- Draw a trend line which best fits your data in exercise number 4.

7. Choose two coordinates that your trend line appears to exactly go through. What is the slope of your trend line?

$$\frac{\text{\_\_\_\_\_ seconds}}{\text{\_\_\_\_\_ students}} = \text{\_\_\_\_\_ sec/student}$$

8. What does the slope represent in the context of this problem? To answer this question, replace the underlined words with values and variables specific to the problem:

"For every <u>*x* unit, *y* unit</u> will (<u>increase/decrease</u>) by <u>slope.</u>"

"For every _____, _____ will _____, by _____."

9. What is the *y*–intercept of your trend line?

_____ seconds

10. What does the y–intercept represent, in the context of this problem? Does this make sense? Explain why or why not.

11. Write the equation of your trend line in slope–intercept form, $y = mx + b$.

12. In this equation, x is the _____ variable. It represents

    _____.

13. In this equation, y is the _____ variable. It represents

    _____.

14. Count how many students are in class: _____ students

15. Predict the time it would take to hand–pulse around the entire class. Follow these steps to assist you in your prediction:

    - Substitute the number of students in class into the equation for your trend line. Be careful to substitute in for the correct variable!
    - You now have an equation with one variable.
    - Solve your equation for the remaining variable.

    The solution to this equation is your prediction for the amount of time it would take a hand–pulse to travel around the entire class.

16. How does this prediction compare with your guess in exercise number 3? How does it compare with your guess in exercise number 5? Which method do you think is the most accurate? Explain.

17. Does this hand–pulse example have a positive, negative, or no correlation? How can you tell?

18. Did our experiment have a strong or weak correlation? How could you justify that your answer is correct?

19. If we all waved hands around the room instead of pulsing hands around the room, would you expect this to have a positive, negative, or no correlation? Why?

20. Using the equation of your trend line, approximately how many students could a hand–pulse travel through in 40 seconds?

21. Approximately how many students are in your school?

22. Approximately how long does your trend line predict it will take for the hand–pulse to travel across this many students?

23. Do you think the trend line makes a good prediction in this case? Why or why not?

24. Compare your prediction with those of your classmates. Whose prediction is the best? Whose is the worst? How do you know?

Hand Pulse Name:ANSWER KEY/TEACHER NOTES
Date:_____

This worksheet is intended to have elements of discovery within. After the hand–pulsing, teacher assistance should be limited to helping students to read and reread directions, as the teacher–instruction will follow the first half of the worksheet. It is very important that all students answer all question prompts. In order to foster discovery, some questions have two or three prompts embedded. Help students to understand that they have not answered an item until all prompts are answered.

1. Have the students stand in a circle starting with two students holding hands, making a circle. Select one student be the timer. When the teacher says go, the timer will time how long it takes for the students to pass a "hand squeezing" pulse around the circle. When the first student squeezes their right hand, the second student will then squeeze their right hand and so forth until the pulse comes back to the first person. Continue to add students to the circle until the desired number of students in the class has been a part of it. Have students record the data in the data table.

2. The time should increase. This is a lead–in to the idea of correlation.

3. Students are expected to look at the data and make a hypothesis, without yet having seen a graph.

4. The graph should have number of students as the *x*–variable and time as the *y*–variable. Points should not be connected.

5. Students are now expected to make a prediction from the graph. This prediction may or may not be different than the prediction for number 3.

6. Students may not come up with the ideal trend line here. In fact, hopefully many lines will be somewhat different. This will lead to varying predictions, especially far outside the data set. It will hopefully then demonstrate the power of using the graphing calculator (tomorrow) for finding the line of best fit, rather than a trend line.

7. To find the slope, students should find the change in *y* and the change in *x* (rise over run) and put them into the blanks. They are then asked to convert into a unit rate, by simplifying.

8. Answers may vary, but should be similar to: "For each student added, the time of the hand–pulsing goes up by (answer to number 7) seconds."

9. The students should find the *y*–coordinate of the point where the trend line crosses the *y*–axis.

10. The *y*–intercept represents the amount of time it would take a pulse to travel through 0 students. Realistically, this answer should be 0. A discussion can be held on the approximate nature of linear regression.

11. Answers may vary, which is important for later.

12. independent; number of students

13. dependent; time

14. If this number is smaller than 16, the teacher may want to consider changing the values in the table back in number 1.

15. The student should correctly substitute his or her answer from number 14 into *x*, using the answer from number 11. The equation should then be correctly solved, and the answer should include units. Note that students may need some assistance with this.

16. The students may think that, because the newest prediction seemed to have the most "math", and because it likely produced some sort of non–integer solution, it is automatically the better one. It is important for the students to realize that they only *approximated* the slope and *y*–intercept for the line which they *approximated* as the trend line. No one prediction is necessarily more accurate than another.

17. This activity most likely has a positive correlation, because as the number of students in the circle increases, the time to complete the pulse should increase.

18. Answers may vary, though the correlation should be relatively strong if one reading is done for each amount of students. Students should be able to justify why they answered the way they did.

19. This should also have a positive correlation, as it is essentially no different from the hand–pulsing.

20. Students should take the equation of the trend line and substitute 40 in for *y* this time. They should then solve for *x*. This may be a difficult exercise for students. Monitor their work closely

21. You may need to help students approximate this.

22. Students should take the equation of the trend line and substitute the answer from exercise 21 in for *x*. They should then solve for *y*.

23. Students will not yet realize that they should not extrapolate the data this far outside the data range. In the coming lessons, the teacher will introduce the concept of 'local linearity', and that sometimes a line of best fit may look appropriate locally, but that

perhaps another model truly fits the data. The error in choosing the wrong model for best fit should be minimal within the data points, but it can be drastic far outside the data.

24. Students will have theories about whose prediction is the best. Most likely, until the students learn to do the line of best fit tomorrow, they will have no way of knowing which prediction came closest to that of the line of best fit. Regardless, since the data point is so far outside the data set, a line may be a poor model, so there truly is no way of knowing for sure whose guess is best.

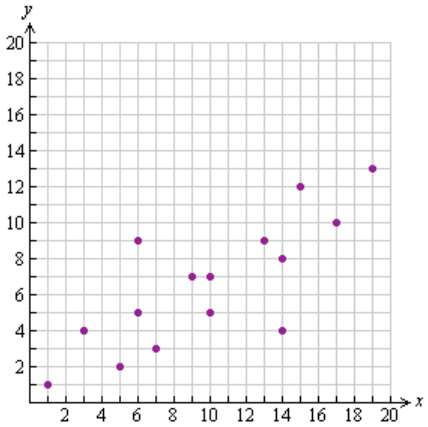
Estimating Lines of Best Fit

Name: _____
Date: _____

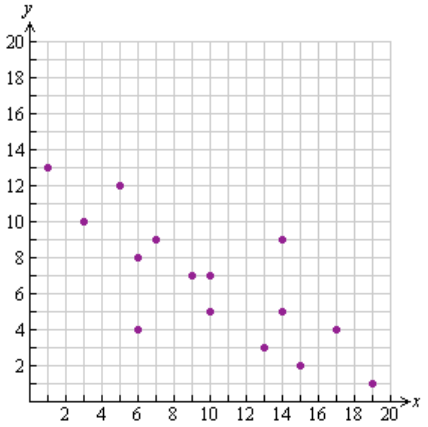
For each scatter plot below:

- Determine the correlation of the scatter plot (positive, negative or no correlation).
- If there is a correlation, sketch a trend line.
- If there is a correlation, estimate the equation of the trend line by hand.

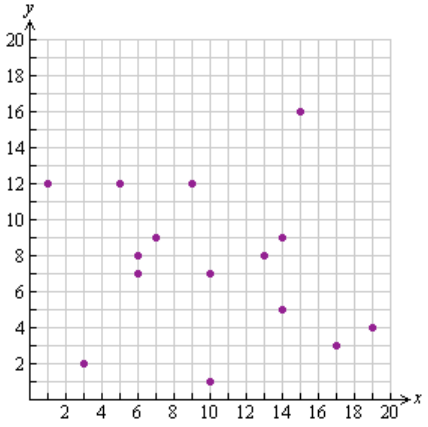
1.



2.



3.



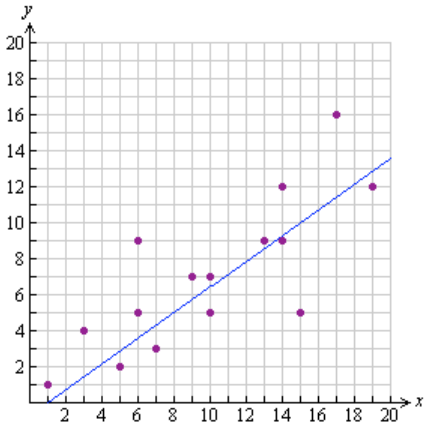
Estimating Lines of Best Fit

Name: __ANSWER KEY__
Date: _____

For each scatter plot below:

- Determine the correlation of the scatter plot (positive, negative or no correlation).
- If there is a correlation, sketch a trend line.
- If there is a correlation, estimate the equation of the trend line by hand.

1.

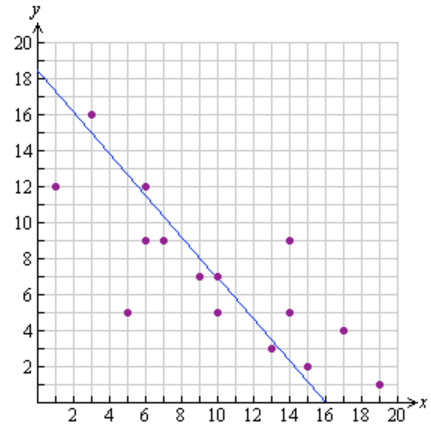


ANSWERS MAY VARY.
These answers are approximate.

Positive Correlation

$y = \frac{5}{7}x - \frac{5}{7}$

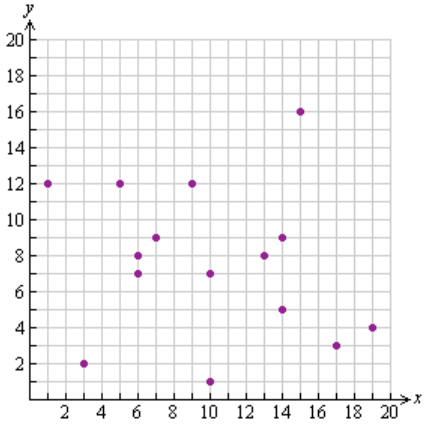
2 .



Negative Correlation

$y = -\frac{15}{13}x + \frac{240}{13}$

3.



No Correlation

Could the Shortest Man in the World See His Shadow? Name: _____
Class: _____

1. In this activity, you will measure each person's height and shadow. Do you predict that this data will have any correlation? _____
   a. If so, what kind of correlation? _____
   b. Explain. _____
      _____

2. Get tape measures from your teacher.

3. Record the time of day: _____

4. Measure the height of everyone in your group in centimeters and record it in the table below.

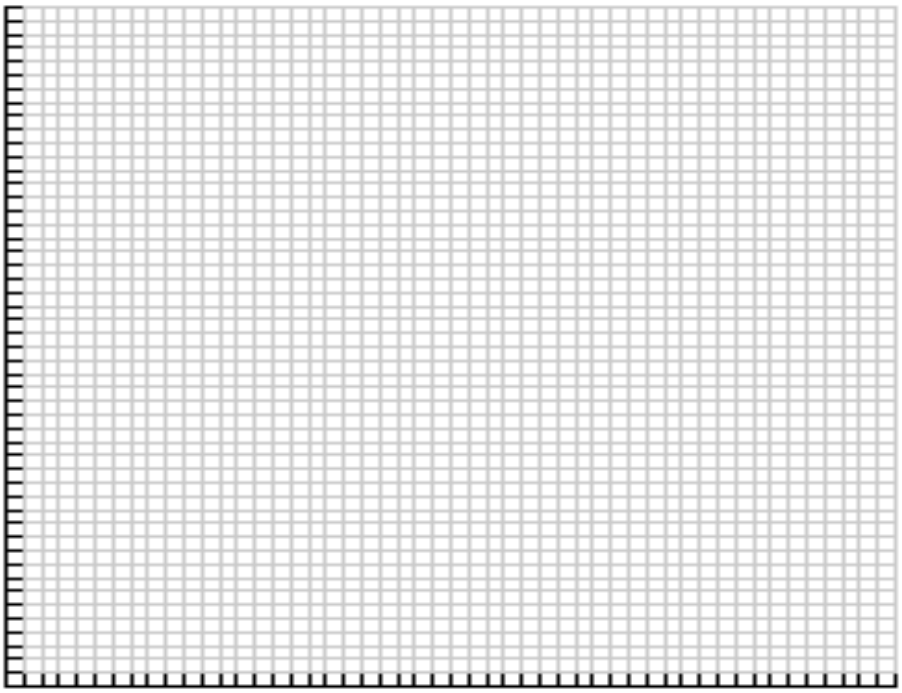
5. Measure the length of everyone in your group's shadow and record it in the table below.

6. Record the heights and shadow lengths from the other groups.

Name	Height	Shadow Length

7. What kind of correlation does the data appear to have? _____
   Is it the same association as you predicted? _____

8. Make a scatter plot of the data using the grid below.



9. What kind of correlation is shown in the scatter plot? _____
   a. Is it the same association as you predicted?_____
   b. What does this kind of association mean in the context of this problem?
      _____
      _____
      _____

10. Draw an estimated trend line on your scatter plot.

11. Estimate the equation of the line of best fit. Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation.

12. What is the slope of your trend line? _____
    What does this slope mean in the context of this problem? _____
    _____
    _____

13. What is the *y*–intercept of your trend line?_____
    What does this *y*–intercept mean in the context of this problem? _____
    _____
    _____

14. If someone is 5'5", or 165.1 cm. tall, about how long would their shadow be?

15. The tallest man in the history of the world was Robert Wadlow. He was 8'11", or 272 cm. tall. Using your equation of best–fit, predict how long his shadow would be.

16. The shortest man in the history of the world was Gul Mohammad. He was 22.5" tall, or 57 cm. Using your equation of best–fit, predict how long his shadow would be.

17. If someone had a shadow that was 190 cm long, how tall would they be, according to your line of best fit?

18. Compare your predictions with the rest of the class. Are they the same or are they different? _____
    If different, how different are they? _____

Complete the remaining questions using the graphing calculator.

19. What is the equation of best fit using the calculator? _____

20. Is the correlation a strong correlation or a weak correlation? _____

21. What is the slope of your line (you may round to the nearest hundredth)? _____
    What does it mean in the context of this problem? _____
    _____
    _____

22. What is the *y*–intercept ? _____
    What does it mean in the context of this problem? _____
    _____
    _____

23. Use your equation to predict how long Robert Wadlow's shadow would be. Do this by hand and use the calculator to check your work.

24. Use your equation to predict how long Gul Mohammad's shadow would be. Do this by hand, and then use your calculator to check your answer.

25. Compare your predictions to your other classmates. Are they the same or are they different? _____
What do you think the reason for that would be? _____
_____
_____

26. Would Gul Mohammad be able to see his shadow? Why or why not? _____
_____
_____

27. Would someone doing this at a different time of day get the same data? Why or why not?_____
_____
_____

At The Pump

Name: _____
Date: _____

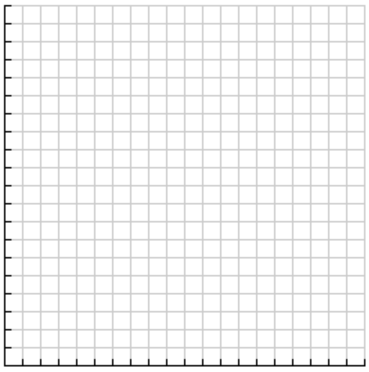
Patrick wants to know how many miles of driving that he gets for every gallon of gas in his car. The number of miles he can drive depends on the amount of gas in his car. The table below shows the number of miles he can drive on different amounts of gallons of gas.

DRIVING DATA

Amount of Gas (gallons)	Distance (miles)
0	0
2	64
4	118
6	181
8	233
10	310

Complete the following:

- Write an equation for a line of best fit. If you choose to draw a graph, use the grid provided.
- What is the slope of your line of best fit? What does the slope mean in the context of this problem?
- Patrick had 20 gallons of gas in his car. According to your line of best fit, how far could he drive? Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation.



At The Pump

Name: _ANSWER KEY___
Date: _____

Patrick wants to know how many miles of driving that he gets for every gallon of gas in his car. The number of miles he can drive depends on the amount of gas in his car. The table below shows the number of miles he can drive on different amounts of gallons of gas.

DRIVING DATA

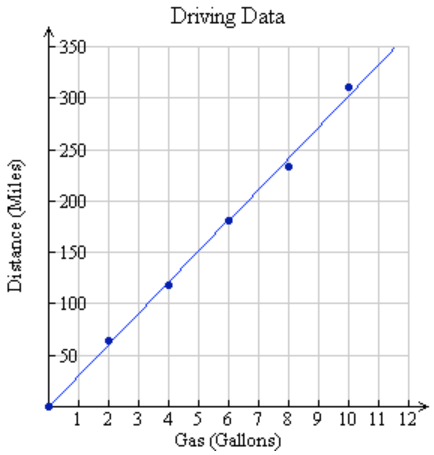
Amount of Gas (gallons)	Distance (miles)
0	0
2	64
4	118
6	181
8	233
10	310

Complete the following:

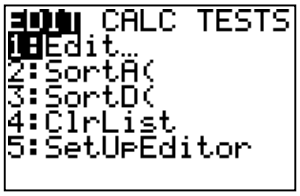
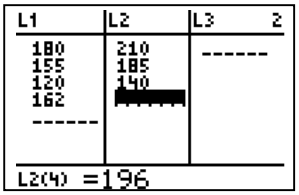
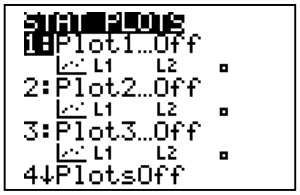
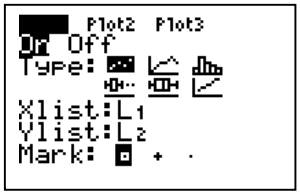
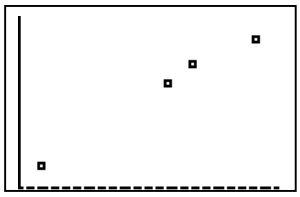
- Write an equation for a line of best fit. If you choose to draw a graph, use the grid provided. $y = 30.29x - 0.43$

- What is the slope of your line of best fit? What does the slope mean in the context of this problem? Slope = 30.29. This slope means that there are 30 miles driven for every gallon of gas.

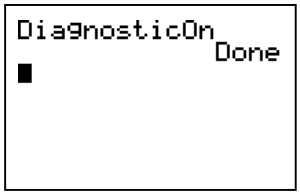
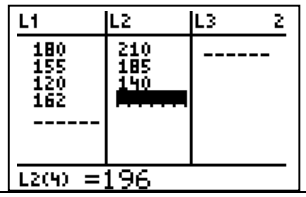
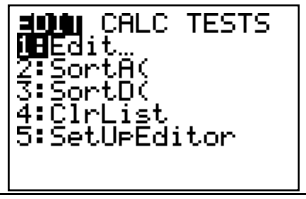
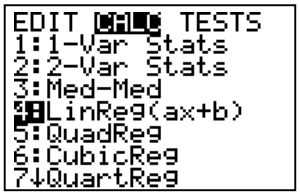
- Patrick had 20 gallons of gas in his car. According to your line of best fit, how far could he drive? Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation. 605.37 miles



To Create A Scatter Plot

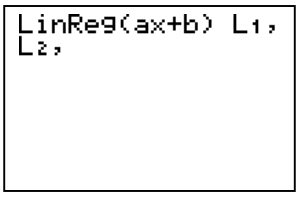
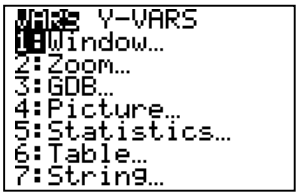
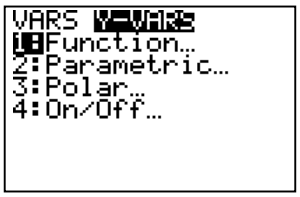
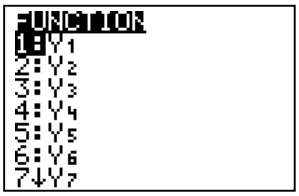
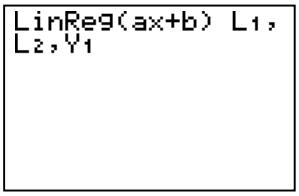
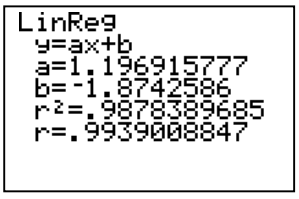
On your calculator, press the STAT button. This will bring up the statistics menu. Press ENTER on number 1 in order to edit your lists.	
Enter in your data, entering the data for the independent variable in L1 and the data for your dependent variable in L2. Make sure to press ENTER after every entry. Make sure that your two lists are the same length when you are done.	
To turn your scatter plot on, press 2ND and then Y=. This will bring you to the STAT PLOTS menu. Your plots should all be off. If they aren't, use option 4 to shut them all off. Then come back to this menu to turn *your* plot on. Press ENTER on Number 1.	
To turn your plot on, move your cursor to highlight the word On on your screen and press ENTER. Make sure also that the picture of the scatter plot is highlighted, and that Xlist and Ylist have L1 and L2, as shown. To type L1, press 2ND 1. To type L2, press 2ND 2.	
To get your scatter plot to fit in your window, hit the ZOOM key, then use your arrow keys to scroll down to number 9:ZoomStat. Hit ENTER.	
Your scatter plot should appear in the window, as shown.	

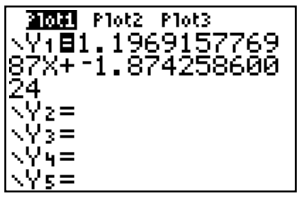
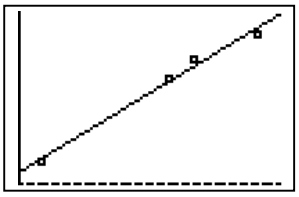
To Find the Line of Best Fit

To make sure that your correlation coefficient will be shown, we have to press 2^{ND} 0 to get to the CATALOG.	
Press the key that has x^{-1} to go to the d's in the list. Then scroll down to DiagnosticOn and hit ENTER.	
Press ENTER again and you will see the word Done. This is how you know that it worked.	
Make sure that your data is in your scatter plot by pressing STAT and hitting ENTER on Edit…	
Press the STAT button.	
Press the right arrow button over to the CALC menu. Use your down arrow buttons to scroll down to number 4:LinReg(ax+b). This is the button for Line of Best Fit, or Linear Regression. Press ENTER	

Some students like to remember these steps with the following mnemonic:

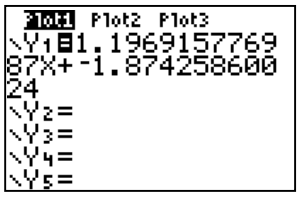
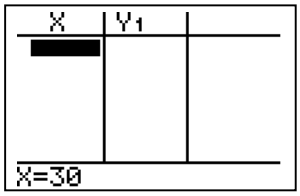
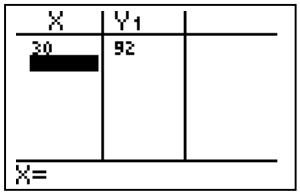
S	**E**	**L**	**L**	**Some**	**Cold**	**Lemonade**
t	d	1	2	t	a	I
a	I			a	l	n
t	t			t	c	R
						e
						g

LinReg(ax+b) will appear on the screen. We need to let the calculator know to use the lists that our data is in. So press 2^{ND} 1 to get L1, then a comma, and 2^{ND} 2 to get L2	
Press the VARS button.	
Push the right arrow button over to the Y–VARS menu. Hit ENTER on 1:Function…	
Hit ENTER on 1:Y1. This will tell your calculator to automatically put your Line of Best Fit into the Y= menu for you so it will automatically graph.	
Your screen should now look like this. Now you should press ENTER to run the program.	
Your Linear Regression screen should look something like this. You can now copy down the equation in the form y=mx+b. Your r= value tells you how close the correlation is. The closer the r is to 1 or –1, the better the fit is.	

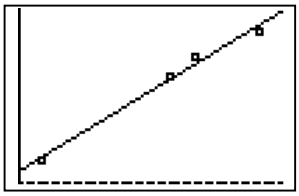
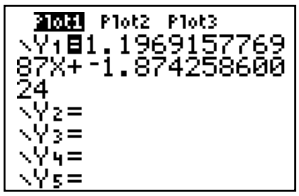
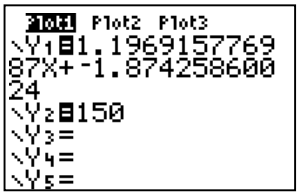
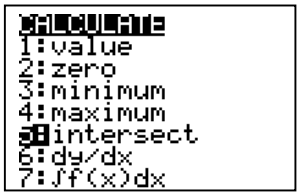
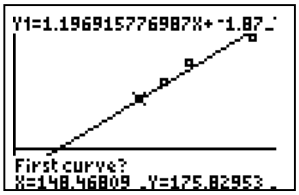
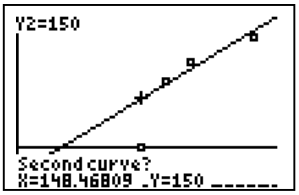
To see your equation in Y=, press the Y= button. Your equation will but put in y=mx+b form into the first Y.	
In order to see your scatter plot and line of best fit, first make sure that your plot is On, and then hit the ZOOM button and go down to 9:ZoomStat and press ENTER	
Now you should see you scatter plot with the line of best fit going through them.	

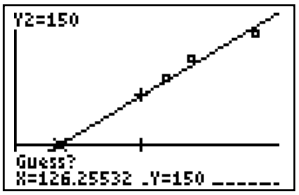
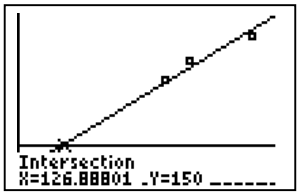
To Predict Values

Predicting y values given an x

Make sure that your equation in is in your Y= menu.	
Hit 2ND WINDOW. This will bring you to TABLE SETUP. Notice the blue word TBLSET over WINDOW. Move the cursor for Indpnt: to Ask. This will allow the calculator to prompt you for an X value.	
Hit 2ND GRAPH. This will bring you to your table representation of your equation. Notice the blue word TABLE over GRAPH. Your cursor will be flashing next to the X=. Enter in whatever X you are looking to predict for.	
Press ENTER. The corresponding Y value will appear in the column under Y1.	

Predicting x values given a y

Make sure that you still have your scatter plot, and line on your screen. Your equation should still be in Y1.	 
Enter the value that you have for your Y value into Y2	
Press 2ND TRACE to get to the CALCULATE menu. Notice the blue CALC above TRACE. Use your down arrow to move down to 5: intersect. Press ENTER	
Both your original line, and the line at the height of your Y value will be graphed on your screen. It will ask you for the First Curve. Your cursor will be a flashing X. Press ENTER.	
It will then ask you for the Second curve. Your flashing X cursor will have automatically moved to the other line. Press ENTER.	

It will then ask you for a Guess. Use your left or right arrows to move your cursor to where the lines meet. Press ENTER.	
The screen will then say Intersection on the bottom. The Y value should be the value that you entered in. The X value will be the value of X when Y is the value that you entered. This would be the answer you were looking for.	

Bring 'Em Home Albert

Name: _____
Date: _____

Born in 1980, Baseball player Albert Pujols is known as one of the best hitters in baseball. As a defensive strategy, opposing pitchers try to walk Pujols rather than allow him to get a hit. Regardless of this, Pujols had the following amount of RBI in his last five years.

Year	2001	2002	2003	2004	2005
RBI	130	127	124	123	117

1. Make a prediction about the type of correlation of this data. Is the correlation positive or negative? Is the correlation weak or strong?

2. Graph the data in a scatter plot using the graphing calculator.

3. Use the graphing calculator to calculate the equation for the line of best fit. Write this equation in the space below.

4. Use your equation to predict the number of RBI's Pujols will have in the 2007 season?

5. Use your equation to predict the first year Pujols will fail to reach at least 100 RBI.

6. Use your equation to predict the number of RBI's Pujols hit in the year 1980? How about 2030? Do these predictions make sense? Explain.

7. Susie argues that the correlation is positive, so Pujols is good at getting RBI. Is this claim correct? Use mathematics to justify your answer.

8. Robert claims that the correlation in this data is negative, which means that Pujols is not good at getting RBI. Is this claim correct? Use mathematics to justify your answer.

Bring 'Em Home Albert

Name: __ANSWER KEY___
Date: _____

Born in 1980, Baseball player Albert Pujols is known as one of the best hitters in baseball. As a defensive strategy, opposing pitchers try to walk Pujols rather than allow him to get a hit. Regardless of this, Pujols had the following amount of RBI in his last five years.

Year	2001	2002	2003	2004	2005
RBI	130	127	124	123	117

1. Make a prediction about the type of correlation of this data. Is the correlation positive or negative? Is the correlation weak or strong?

   negative, (very) strong

2. Graph the data in a scatter plot using the graphing calculator.

   See student work.

3. Use the graphing calculator to calculate the equation for the line of best fit. Write this equation in the space below.

   $y = -3x + 6133.2$ *or* $y = -3x + 133.2$, depending on whether 2001, 2002, 2003, etc. or 1, 2, 3, etc. are used for years.

4. Use your equation to predict the number of RBI's Pujols will have in the 2007 season?

   112 RBI's

5. Use your equation to predict the first year Pujols will fail to reach at least 100 RBI.

   Year 2012

6. Use your equation to predict the number of RBI's Pujols hit in the year 1980? How about 2030? Do these predictions make sense? Explain.

   For 1980, he would have had 193 RBI's. However, since he was born in 1980, this answer does not make sense. In 2030, he will have 43 RBI's. However, he will be 50 years old and may be retired by then.

7. Susie argues that the correlation is positive, so Pujols is good at getting RBI. Is this claim correct? Use mathematics to justify your answer.

   Susie is wrong because the correlation is not positive.

8. Robert claims that the correlation in this data is negative, which means that Pujols is not good at getting RBI. Is this claim correct? Use mathematics to justify your answer.

   Robert is correct to state that the correlation is positive. Note, however, that Robert is incorrect because a negative correlation does not necessarily signify poor performance. It only suggests that the performance is decreasing from whichever point it started.

Olympic Gold Medals

Men's 100m

Year	Time (sec)
1896	12.0
1900	11.0
1904	11.0
1908	10.8
1912	10.8
1920	10.8
1924	10.6
1928	10.8
1932	10.38
1936	10.3
1948	10.3
1952	10.79
1956	10.62
1960	10.32
1964	10.06
1968	9.95
1972	10.14
1976	10.06
1980	10.25
1984	9.99
1988	9.92
1992	9.96
1996	9.84
2000	9.87
2004	9.85

Women's 100m

Year	Time (sec)
1896	——
1900	——
1904	——
1908	——
1912	——
1920	——
1924	——
1928	12.2
1932	11.9
1936	11.5
1948	11.9
1952	11.65
1956	11.82
1960	11.18
1964	11.49
1968	11.08
1972	11.07
1976	11.08
1980	11.06
1984	10.97
1988	10.54
1992	10.82
1996	10.94
2000	10.75
2004	10.93

http://www.hickoksports.com/history/olmtandf.shtml#m100

Olympic Gold Medals

Men's High Jump

Year	Height (m)
1896	1.81
1900	1.9
1904	1.8
1908	1.9
1912	1.93
1920	1.935
1924	1.98
1928	1.94
1932	1.97
1936	2.03
1948	1.98
1952	2.04
1956	2.12
1960	2.16
1964	2.18
1968	2.24
1972	2.23
1976	2.25
1980	2.36
1984	2.35
1988	2.38
1992	2.34
1996	2.39
2000	2.35
2004	2.36

Women's High Jump

Year	Height (m)
1896	——
1900	——
1904	——
1908	——
1912	——
1920	——
1924	——
1928	1.59
1932	1.657
1936	1.6
1948	1.68
1952	1.67
1956	1.76
1960	1.85
1964	1.9
1968	1.82
1972	1.91
1976	1.93
1980	1.97
1984	1.91
1988	2.03
1992	2.02
1996	2.05
2000	2.01
2004	2.06

http://www.hickoksports.com/history/olwtandf.shtml#highj

Olympic Gold Medals

Name :_____
Date: _____

1. Observe the data for the Olympic gold medals in the Men's and Women's 100m dash.

   - Notice that certain years are missing from both tables. Why do you think these years are missing?

   - The Women's data table is missing times for the years 1896 - 1924. Why do you think the data is missing for these years?

2. Choose either the Men's 100m dash or the Women's 100m dash to answer the following:

   - I chose the _____ 100m dash.

   - Use your calculator to calculate the equation for the line of best fit.

   - Is there a correlation between the year and the gold medal 100m dash time? If so, what kind of correlation does this data have?

   - Does this association surprise you? Why or why not?

   - What is the slope of your line (rounded to the nearest hundredth)? What does it mean in the context of this problem?

   - What is your *y*–intercept (rounded to the nearest hundredth)? What does it mean in the context of this problem?

   - How do you think human characteristics have influenced the trend?

   - Use your equation for the line of best fit to predict the Olympic record for the 100m dash in 2040? in 2080? Are these realistic predictions? Why or why not?

   - If someone asked you if we can predict future gold medal 100m dash times based upon the year, what would you say? Use mathematics to justify your answer.

3., Observe the data for the Olympic gold medals in the Men's and Women's high jump. Choose either the Men's high jump or Women's high jump to answer the following:

- I chose the _____ high jump.
- Use your calculator to calculate the equation for the line of best fit.
- Is there a correlation between the year and the high jump height? If so, what kind of correlation does this data have?
- Does this association surprise you? Why or why not?
- What is the slope of your line (rounded to the nearest hundredth)? What does it mean in the context of this problem?
- What is your *y*–intercept (rounded to the nearest hundredth)? What does it mean in the context of this problem?
- How do you think human characteristics have influenced the trend?
- Use your equation for the line of best fit to predict the Olympic record for the high jump in 2040? In 2080? Are these realistic predictions? Why or why not?
- If someone asked you if we can predict future gold medal high jump heights based upon the year, what would you say? Use mathematics to justify your answer.

Olympic Gold Medals

Name :__ANSWER KEY___
Date: _____

1. Observe the data for the Olympic gold medals in the Men's and Women's 100m dash.

   - Notice that certain years are missing from both tables. Why do you think these years are missing?
     The Olympics were cancelled due to World War I and World War II.

   - The Women's data table is missing times for the years 1896 - 1924. Why do you think the data is missing for these years?
     Women were not allowed into this competition until 1928.

2. Choose either the Men's 100m dash or the Women's 100m dash to answer the following:

   - I chose the _____ 100m dash.

   - Use your calculator to calculate the equation for the line of best fit.
     $y_{\text{men}} = -0.013x + 11.096$ $y_{\text{women}} = -0.018x + 12.489$

   - Is there a correlation between the year and the gold medal 100m dash time? If so, what kind of correlation does this data have?
     Yes, negative correlation

   - Does this association surprise you? Why or why not?
     Answers may vary.

   - What is the slope of your line (rounded to the nearest hundredth)? What does it mean in the context of this problem?
     $m_{\text{men}} = -0.01$ $m_{\text{women}} = -0.02$ Every year the 100m dash time will drop by this many seconds

   - What is your *y*–intercept (rounded to the nearest hundredth)? What does it mean in the context of this problem?
     $b_{\text{men}} = 11.096$ $b_{\text{women}} = 12.489$. These are the approximate times of the 100m dash for the year 1900.

   - Use your equation for the line of best fit to predict the Olympic record for the 100m dash in 2040? in 2080? Are these realistic predictions? Why or why not?

     Men's 100m dash: y=-0.01x+11.096

y=-.0.01(144)+11.096
y=9.696
y=-0.01(184)+11.096
y=9.256
Women's 100m dash: y=-.02x+12.489
y=-.02(112)+12.489
y=10.249
y=-.02(152)+12.489
y=9.449

- How do you think human characteristics have influenced the trend?
  Answers may vary, such as health, technology, physics, etc.

- If someone asked you if we can predict future gold medal 100m dash times based upon the year, what would you say? Use mathematics to justify your answer.
  The further into the future the predictions go, the less accurate they become because they are further away from the data set.

3. Observe the data for the Olympic gold medals in the Men's and Women's high jump. Choose either the Men's high jump or Women's high jump to answer the following:

   - I chose the _____ high jump.
     Answers will vary

   - Use your calculator to calculate the equation for the line of best fit.
     Men's High Jump: y=.0056x+1.80
     Women's High Jump: y=.0066x+1.59

   - Is there a correlation between the year and the high jump height? If so, what kind of correlation does this data have?
     There is a positive correlation

   - Does this association surprise you? Why or why not?
     Answers may vary

   - What is the slope of your line (rounded to the nearest hundredth)..What does it mean in the context of this problem?
     Men's: m=.006 Women's: .007
     It means that the men's high jump increased .006 meters per year and the women's high jump increased .007 meters per year

- What is your *y*–intercept (rounded to the nearest hundredth)? What does it mean in the context of this problem?

  Men's: b=1.80 Women's: b=1.59

  This means that the men's jump height started at 1.80 meters and the women's jump height started at 1.59 meters

- How do you think human characteristics have influenced the trend?

- Use your equation for the line of best fit to predict the Olympic record for the high jump in 2040? in 2080? Are these realistic predictions? Why or why not?

  Men's High Jump: $y=.0056x+1.80$

  $y=.0056(144)+1.80$

  $y=2.6064$

  $y=.0056(184)+180$

  $y=2.8304$

  Women's High Jump: $y=.0066x+1.59$

  $y=.0066(112)+1.59$

  $y=2.3292$

  $y=.0066(152)+1.59$

  $y=2.5932$

- If someone asked you if we can predict future gold medal high jump heights based upon the year, what would you say? Use mathematics to justify your answer.

  The further into the future the predictions go, the less accurate they become because they are further away from the data set.

Name: _____
Date: _____

Air Pollution and Asthma

State	NO_x (nitrus oxide, tons)	# of Power Plants	# of Asthma Attacks
MD	71,454	7	17,325
PA	200,932	25	35,405
WV	225,857	13	7,293
OH	368,735	24	39,703
IL	174,247	25	33,986
IN	281,117	21	21,532
KY	198,541	19	16,440
VA	78,438	15	23,736
AR	42,131	5	8,327
NC	145,705	18	27,418
MS	44,349	3	7,341
AL	161, 563	9	13,480

2002 Data Source: http://www.cleartheair.org/dirtypower/map.html

1. Use your calculator to find the equation for the line of best fit comparing NO_x and the number of asthma attacks.

2. Use your calculator to find the equation for the line of best fit comparing the number of power plants and NO_x

3. Is there a correlation between NO_x and the number of asthma attacks? If so, what kind of correlation does this data have? Does this correlation surprise you? Why or why not?

4. What is the slope of your line (rounded to the nearest hundredth)? What does it mean in the context of this problem?

5. Is there a correlation between the number of power plants and NO_x? If so, what kind of correlation does this data have? Does this correlation surprise you? Why or why not?

6. What is the slope of your line (rounded to the nearest hundredth)? What does it mean in the context of this problem?

Air Pollution and Asthma Name: ___ANSWER KEY__
Date: _____

Air Pollution and Asthma

State	NO_x (nitrus oxide, tons)	# of Power Plants	# of Asthma Attacks
MD	71,454	7	17,325
PA	200,932	25	35,405
WV	225,857	13	7,293
OH	368,735	24	39,703
IL	174,247	25	33,986
IN	281,117	21	21,532
KY	198,541	19	16,440
VA	78,438	15	23,736
AR	42,131	5	8,327
NC	145,705	18	27,418
MS	44,349	3	7,341
AL	161, 563	9	13,480

2002 Data Source: http://www.cleartheair.org/dirtypower/map.html

1. Use your calculator to find the equation for the line of bes fit comparing NO_x and the number of asthma attacks.
   $y = 0.063x + 10586.742$

2. Use your calculator to find the equation for the line of best fit comparing the number of power plants and NO_x
   $y = 9261.1x + 24085.62$

3. Is there a correlation between NO_x and the number of asthma attacks? If so, what kind of correlation does this data have? Does this correlation surprise you? Why or why not?
   Yes, positive, answers may vary

4. What is the slope of your line (rounded to the nearest hundredth)? What does it mean in the context of this problem?
   Slope is 0.06, which means for each ton of nitrus oxide produced, the number of asthma attacks goes up by 0.06

5. Is there a correlation between the number of power plants and NO_x? If so, what kind of correlation does this data have? Does this correlation surprise you? Why or why not?
   Yes, positive, answers may vary

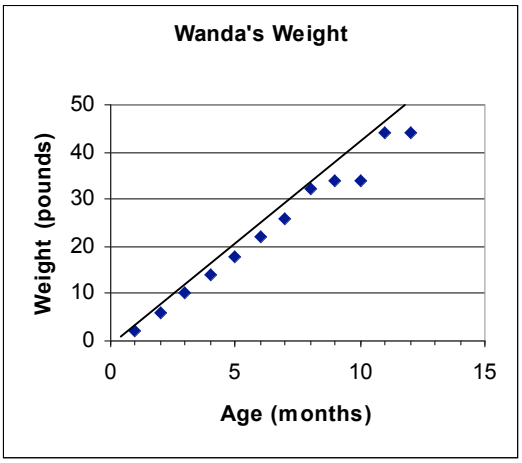
6. What is the slope of your line (rounded to the nearest hundredth)? What does it mean in the context of this problem?
   Slope is 9261.1, which means for every additional power plant, the amount of nitrus oxide released into the air increases by 9261 tons.

Linear Regression
Summative Assessment

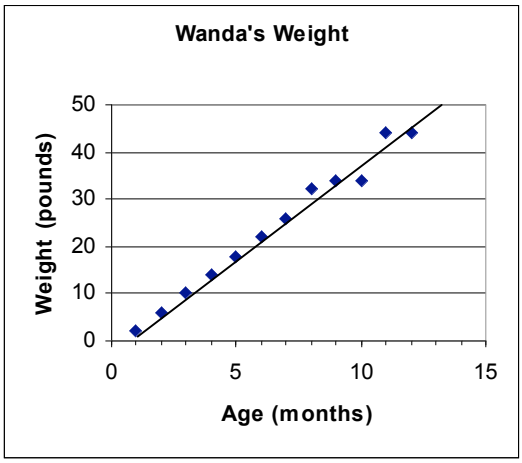
Name: _____
Class: _____

1. David keeps track of the weight of his dog, Wanda, over the course of a year. Which of the following shows data with a correct line of best fit?

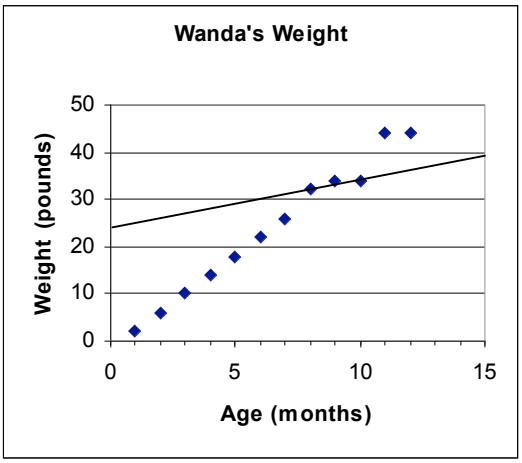
A



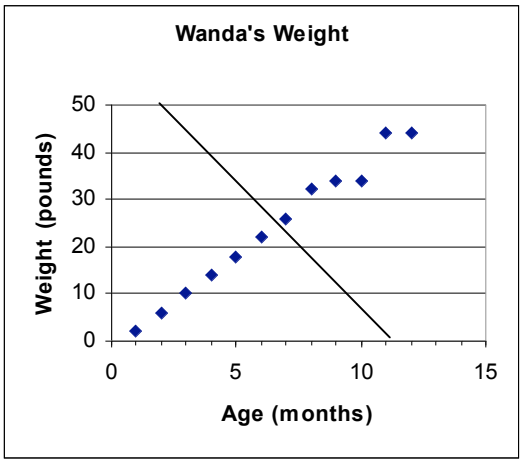
B



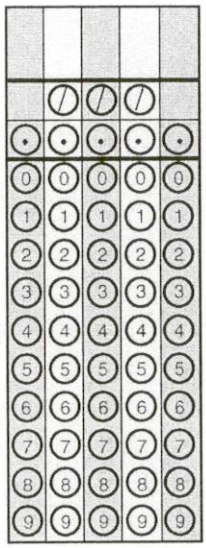
C

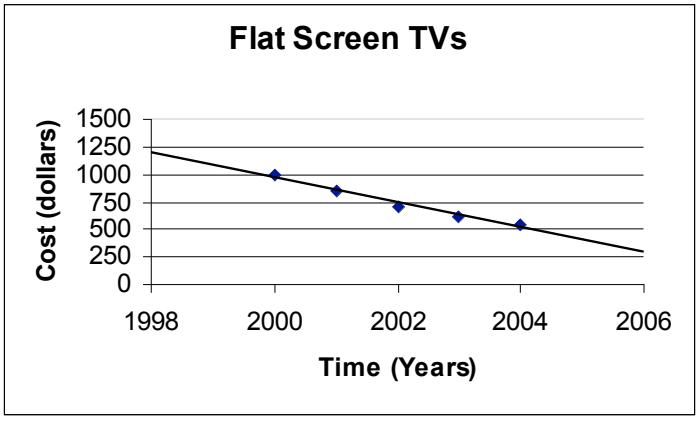


D



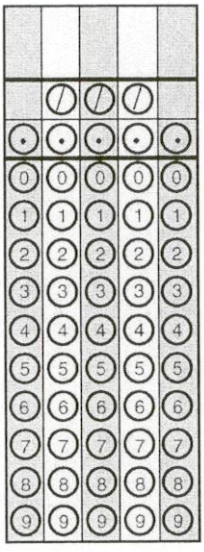
2. The cost of flat screen televisions are recorded on the graph shown.





According to the line of best fit, approximately how much would a flat screen television cost in 2006?

3. Bill always deletes his financial records that are more than 7 years old. The following table records how much money Bill had in his savings account at the end of each year, over the last 7 years.

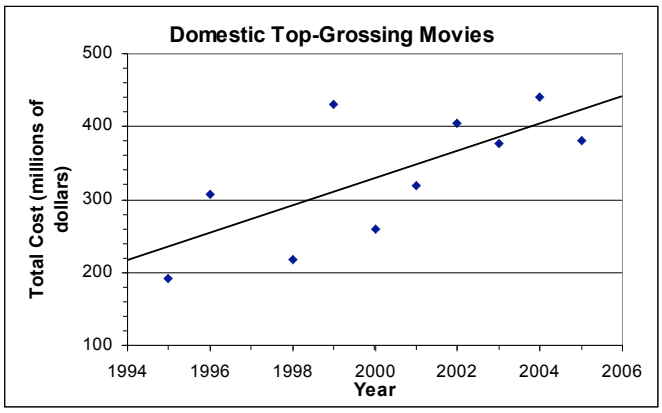


Year	1999	2000	2001	2002	2003	2004	2005
Balance	$1,533	$1,745	$1,976	$2,212	$2,301	$2,625	$2,852

According to the line of best fit, what was the beginning balance in Bill's account at the end of 1995?

4. Jim was analyzing the domestic top–grossing movies of the past ten years. Unfortunately, he could not find any data for *Titanic*.

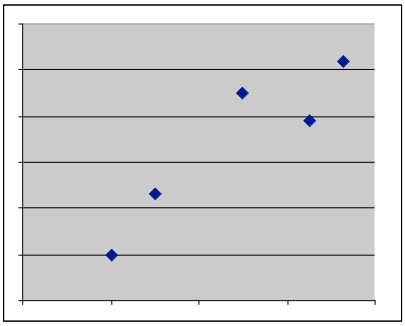
Top Grossing Title	Year	Total Gross (millions of dollars)
Toy Story	1995	192
Independence Day	1996	306
Titanic	1997	
Saving Private Ryan	1998	217
The Phantom Menace	1999	431
The Grinch	2000	260
Harry Potter	2001	318
Spider–Man	2002	404
Return of the King	2003	377
Shrek 2	2004	441
Revenge of the Sith	2005	380



Based on his line of best fit, what should be his prediction for the domestic gross of *Titanic*?

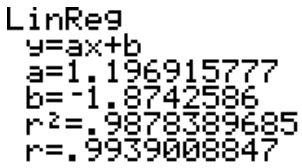
A) $270 million
B) $200 million
C) $235 million
D) It would be improper to use the line of best fit to make a prediction this far inside the data set.

5. What kind of association does the following graph show?



A) Positive association
B) Negative association
C) No association
D) None of the above

6. A linear regression was performed on a set of data . According to this screen shot, is there a good correlation between the x and y variables?



A) Yes, because the slope is close to 1
B) No, because the slope is close to 1
C) Yes, because the correlation coefficient is close to 1
D) No, because the correlation coefficient is close to 1

7. Jeremy is buying a car. He does some research on what happens to the car's value over time. This can be represented by the equation $y = -1000x + 25000$, where x represents the number of years and y represents the value of the car. What does the slope mean in the context of this situation?

A) The car increases in value $25,000 every year
B) The car decreases in value $1000 every year
C) The car originally is worth $1000
D) The car originally is worth $25,000

8. Callie is saving money. The equation of best fit to represent how much money she saves each year is $y = 1200x + 500$, where x represents the number of years and y represents the amount of money saved. What does the y – intercept mean in the context of the situation?

A) She started with $1200
B) She saves $1200 per year
C) She started with $500
D) She saves $500 per year

9. A positive slope means which of the following?

A) As the x – value decreases the y – value increases
B) As the x – value increases the y – value decreases
C) As the x – value increases the y – value is constant
D) As the x – value increases the y – value increases

10. What can be said about the slope of a graph that has a negative association?

A) The slope is positive
B) The slope is negative
C) The slope is 0
D) There is no slope

11. A shoe company wants to determine how quickly the sole of its shoes wears down with average use. Let x represent the number of months the shoe was worn. Let y represent the thickness of the sole in millimeters. An equation for a line of best fit is shown below.

$$y = -x + 16$$

- What is the slope of this line of best fit? What does the slope mean in the context of this problem?

- What is the y–intercept of this line of best fit? What does the y–intercept mean in the context of this problem?

- Jocelyn will need to replace her shoes when they have 2 millimeters of tread left. According to the line of best fit, for how many months can Jocelyn wear her shoes before she needs to replace them? Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation.

12. John and his friends compared notes on how many points were scored on a video game and how much time they had spent playing. They came up with the following table:

Time (min)	25	35	40	55	65	80	95
Score	50	60	75	95	100	90	100

- Write the equation for a line of best fit.

- Explain what the slope and y–intercept of your equation represent in this context.

- Predict how many points they would score if they spent two hours playing the video game. Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation.

Linear Regression Assessment Answer Key

1. B
2. Acceptable Range: 250–270
3. 667.2
4. A
5. A
6. C
7. B
8. C
9. D
10. B

11.

- The slope of the line of best fit is negative 1. This means that the sole of the shoe wears down one millimeter every month.

- The y–intercept of the line is 16. This means that the shoe starts with 16 millimeters on the sole.

- $y = -x + 16$

  $2 = -x + 16$

  $-14 = -x$

  $14 = x$

  Jocelyn will need to replace her shoes in 14 months.

- 12.
  - $y = .68x + 43.02$

  - The slope is .68 and it means that score increases by approximately .68 points per minute played. The *y*–intercept is 43.02. This means that you start with approximately 43.02 points.

  - $y = .68x + 43.02$

    $y = .68(120) + 43.02$

    $y = 81.6 + 43.02$

    $y = 124.62$

    John and his friends would score approximately 125 points if they were to play for 2 hours, or 120 minutes.

Authors:

Evelyn Kytic
Pocono Mountain East High School
Monroe County

Joseph Sutton
Calvert High School
Calvert County

Emily Yanisko
Parkdale High School
Prince George's County